# A SORGHUM NAC GENE IS ASSOCIATED WITH VARIATION IN VASCULAR DEVELOPMENT, BIOMASS PROPERTIES, AND YIELD POTENTIAL

BY

JINGNU XIA

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Masters of Sciences in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Adviser:

Associate Professor Patrick Brown (Chair)
Professor Stephen Moose
Professor Matthew Hudson

# ABSTRACT

*Sorghum bicolor* is an important cereal, forage and energy crop adapted to semiarid areas. Primary goals for crop breeding include maximizing sugar and grain yield while maintaining the rigidity and strength of the stalk. Vascular architecture influences not only stalk strength and rigidity, but also water, nutrient and sugar transport, which ultimately determine grain yield. The building blocks of the vascular system are secondary cell walls, which make up the majority of plant biomass. Understanding the gene regulatory network underlying secondary cell wall synthesis can improve the understanding of vascular development and yield and ultimately benefit future cereal breeding. This thesis describes the mapping of the *Dry Midrib (D)* locus in sorghum, which controls a phenotypic difference between juicy green *(dd)* and dry white *(DD)* vascular tissue. The first chapter describes fine-mapping and GWAS using green/white midrib color as the mapping phenotype. Near isogenic lines (NILs) were created, and flanking markers and genotyping-by-sequencing (GBS) were applied to $BC_3$ and $BC_2F_2$ individuals to reduce the region of interest to a 38 kb interval containing 4 candidate genes. One of these genes is a NAC gene containing a stop codon mutation in the mutant *(dd)* allele. A preliminary analysis of transcriptome differences in midrib tissue of *dd* and *DD* NILs is also described. The second chapter compares differential gene expression analysis in midrib and stalk tissue of *dd* and *DD* NILs. A glycosyl hydrolase gene is differentially expressed between *dd* and *DD* NILs in midrib, stalk, and midrib + stalk analyses. Gene ontology analysis in all three sets showed significant enrichment of hydrolase activities on O-glycosal compounds and glycosal bonds in genes differentially-expressed between *dd* and *DD*. Overall, these results provide valuable hints to fine-tune sorghum breeding in the future.

# ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Patrick Brown for enduring my seemingly endless drafts, and my two committee members Dr. Stephen Moose and Dr. Matthew Hudson for providing excellent advice and feedback. I owe the success of this project to the collaborators who helped greatly on the bench and in the field. I would like to appreciate Dr. Markus Pauly and Dr. Yunjun Zhao collecting all cell wall biocompositional data. I also would like to appreciate Dr. Payne Burks, who collected the lodging data. Finally, I appreciate my family and loved ones for supporting me in my life and career.

This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant no DE-SC0012400.

# TABLE OF CONTENTS

# Literature Review

*Sorghum bicolor*, a $C_4$ grass, originates from the steppes and savannas of sub-Saharan Africa. Currently, sorghum is a staple grain for more than 500 million people, which makes it the fifth most important cereal after rice, wheat, maize, and barley. Sorghum has a wide spectrum of uses, ranging from grain consumption by humans to feedstocks for distillation, animal feed and bioenergy. According to USDA, in 2016 the world production of sorghum was 62.8 million tons while the consumption of sorghum was 63.1 million tons ("Grain: World Markets and Trade | USDA Foreign Agricultural Service," n.d.; Jihong Wu et al., 2015; More et al., 2016; Smith and Frederiksen, 2000). Sorghum is a resilient crop that can tolerate low precipitation environments. This makes it a promising crop for semi-arid areas where it won't compete with other major staple crops.

Sorghum has great genetic variation in carbon partitioning, sugar storage, cell wall mass, perenniality and tillering. There are three major types of sorghums -- grain, sweet and biomass sorghum. Grain sorghums are used for grain consumption. Sweet sorghums are used in both syrup production and the bioenergy industry. Biomass sorghums are mostly photoperiod-sensitive tropical lines, which have indeterminate growth and are used as forage and bioenergy feedstocks (Godin et al., 2016; Paterson et al., 2009).

Plant architecture affects adaptation to the environment. Inflorescence and leaf architecture have been heavily studied in grasses (*Poaceae*), but vascular system (xylem and phloem) architecture has received less attention. The vascular system not only transports water, nutrients and sugars, but also provides mechanical support and rigidity. The major building blocks for all plant organs are cell walls. The primary contents of the cell wall are polymers and

fibers, such as lignin, cellulose, hemicellulose, and. Differential composition of polymers and fibers will determine characteristics of the cell wall and vascular system and will consequently affect whole plant physiology. Cell wall synthesis is controlled by a complex gene network. Understanding this network can assist with understanding genetic control of vascular system architecture and potentially improve critical agronomical traits (Peiffer et al., 2013; Zhong and Ye, 2015).

*Cell Wall Physiology*

Higher plants contain primary and secondary cell walls. Primary cell walls are produced in all cells, and include cellulose, hemicellulose and pectin. Secondary cell walls are produced after cell growth has ceased, and are deposited in great abundance in tracheary elements, fibers and other sclerenchymatous cells in the vascular system. Secondary cell walls make up the majority of plant biomass. and include cellulose, hemicellulose and lignin. Thick secondary cell walls surround thinner primary cell walls and determine the cell's final size. Cellulose is synthesized in the plasma membrane. Hemicellulose and pectin are synthesized in the Golgi apparatus and transported to the cell wall surface by vesicles. Hemicellulose and pectin combines together and crosslink cellulose fibers. Monolignols are synthesized in the ER and assembled into lignin polymers in the apoplast. Studies have shown that reducing lignin and cellulose can increase saccharification yield but result in pendant stems and brittle culms (Cosgrove, 2005; Kumar et al., 2016; Vargas et al., 2016; Wang et al., 2016).

During primary cell wall growth, the cell wall is a thin, strong and flexible layer. Although the layer is thin, it has a considerable strength that holds the shape of the cell. The outer layer mostly consists of cellulose and microfibrils. The expansion of the primary cell wall

occurs through action of various polymers and fibers. Pectin forms hydrated gels, which push cellulose microfibrils away to the sides and allow the addition of new cellulose polysaccharides. When the secondary cell wall is formed the cell is no longer extendable (Cosgrove, 2005).

*NAC Transcription Factors (TFs)*

The synthesis of cellulose, hemicellulose, pectin and lignin involves a complicated network of genes. NAC and MYB are the two primary classes of TFs involved in the cell wall transcriptional network. NAC TFs are named after the collection of NAM, ATAF and CUC TFs, are specific to plants, and are often cited as "master regulators" of secondary cell wall synthesis (Vargas et al., 2016; Zhong and Ye, 2015) All NAC TFs are characterized by a highly conserved NAC domain at the N terminus. The assembly of NAC sub-domains determines NAC subfamilies and protein functions. NAC genes within a subfamily have highly conserved function (Nuruzzaman et al., 2012a; Peng et al., 2015a).

Accumulated evidence shows that NAC TFs have multiple functions in plant growth and development, including stress response, hormone signaling, and organ development. (Peng et al., 2015a; Xie et al., 2000; Zhong et al., 2006). There are several subfamilies of NAC genes that are involved in stress response. Under drought and salinity stress, transcriptional analysis suggests that 5 OsNAC genes are subject to ABA and JA elevation and target downstream stress response genes (Kazuo Nakashima et al., 2012). NAC TFs have also been reported to involve in cell wall synthesis. Double knockout of NST1 (NAC SECONDARY WALL THICKENING PROMOTING FACTOR1) and NST3/SND1 (SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN1) in Arabidopsis prevents formation of a secondary cell wall. In the vascular vessels, VASCULAR-RELATED NAC-DOMAIN6 (VND6) and VND7 have the dual

role of regulating secondary cell wall development and cell senescence. VND6, VND7 and SND1 are classified in the same NAC subfamily and control similar downstream genes, but cellular senescence in the vascular system can only be triggered by VND6 and VND7. In Arabidopsis, NAC1 activates two auxin response genes, DBP and AIR3, and consequently affects lateral root development. The CUC1 subfamily of genes regulates the development and separation of the embryo, vegetative and floral organs (Mallory et al., 2004; Xie et al., 2000; Zhong et al., 2006).

# Chapter One: A sorghum NAC gene is associated with variation in vascular development, biomass properties, and yield potential


## SUMMARY

*Sorghum bicolor* is a $C_4$ grass widely cultivated for grain, forage, sugar and biomass. The sorghum *Dry Stalk (D)* locus controls a qualitative difference between juicy green *(dd)* and dry white *(D-)* stalks and midribs, and co-localizes with a quantitative trait locus for sugar yield. Here, we apply fine-mapping and GWAS to identify a candidate gene underlying *D*, and use nearly-isogenic lines to characterize the transcriptional, compositional, and agronomic effects of variation at the *D* locus.

The *D* locus was fine-mapped to a 38 kb interval containing four genes. One of these genes is a NAC transcription factor that contains a stop codon in the NAC domain in the recessive *(dd)* parent. Allelic variation at *D* affects grain and sugar yield, biomass composition, and vascular anatomy in nearly-isogenic lines. Green midrib (*dd)* NILs show reductions in lignin and the number of vascular bundles in stalk tissue, but produce higher sugar and grain yields under well-watered field conditions. Increased yield potential in *dd* NILs is associated with increased stalk mass and moisture, higher biomass digestibility, and an extended period of grain filling. Transcriptome profiling of midrib tissue at the 4-6 leaf stages, when NILs first become phenotypically distinct, reveals that *dd* NILs have increased expression of a miniature zinc finger (MIF) gene. MIF genes dimerize with and suppress zinc finger homeodomain (ZF-HD) transcription factors, and a ZF-HD gene is associated with midrib color variation in a GWAS analysis across 1694 diverse sorghum inbreds.

A premature stop codon in a NAC gene is the most likely candidate polymorphism underlying the sorghum *D* locus. *dd* NILs produce more grain, sugar, and biomass under well-watered field conditions. Our results suggest that interaction between NAC, MIF, and MIF-targeted ZF-HD transcription factors may regulate vascular development and agronomic potential in cereals.

## METHODS

*Plant material and controlled growth conditions*

A dominant white midrib allele *(D)* was introduced into a genetic male sterile *(ms3/ms3)* version of Tx623, which carries a recessive green midrib allele *(d),* through four generations of backcrossing. Homozygous *D/D-Ms3/Ms3* and *d/d-Ms3/Ms3* seed stocks were derived from a single $BC_4$ plant by selfing. Three replications of greenhouse plantings were performed on 1/7/16, 2/16/16 and 3/16/16, planted 2 seeds per cell in 6 by 8 flats and thinning to 1 plant per cell at 7 days after planting. Midrib color changes were recorded using a MiScope (Zarbeco, USA) at V4, V6, and V8, and midrib tissue for RNAseq was sampled at V4 and V6 into liquid N. Total number of vascular bundles in cross-sections of young expanding *DD* and *dd* internodes were counted at the 6-leaf stage; staging was performed by selecting the internode from each plant that was closest to 5 mm in diameter.

*Microscopy*

Stalk tissue was sampled from field-grown young (~5 mm diameter) internodes at the 8-leaf stage. Samples were submerged in 1X PBS solution, held under vacuum for 10 minutes, and transferred to a Leica ASP300 (Leica, Germany) for 72-hour tissue processing. Fixed tissues were embedded using a Leica EG1150 paraffin embedding station and cold plate, and paraffin blocks were sectioned to 15 μm using a Leica RM2250 microtome. Sections were mounted on

glass slides and stained with fast-green and safranin following the protocol of Ma et al (Ma et al., 1993). All light microscopy measurements were made with a NanoZoomer Digital Pathology System (Hamamatsu, Japan) using 20X magnification.

*Field experiments*

For sugar yield measurements, paired rows of NILs were planted in six replicates at a single location in Urbana, IL in summer 2014, and phenotyped as previously described (Burks et al., 2015). For grain and stalk dry matter and moisture measurements, paired 2-row plots of NILs were planted in four replicates at each of two locations (Urbana and Savoy, IL), and samples were pooled from 2-6 individual representative plants per row for each time point. Immature grain weight was estimated by clipping individual panicle branches from the rachis, stalk weight was measured after stripping off leaf blades, and moisture was estimated by weighing samples before and after oven-drying at 45°C for five days. Midrib color in the GWAS panel was phenotyped as a binary trait (green vs. white) in the youngest leaf at ~45 days after planting. All field experiments were machine planted in rows 10' long with 30" row spacing. Seeds were treated with Apron fungicide (Syngenta, USA) and Concep II seed safener (Syngenta, USA) before planting, and weeds were controlled through use of a pre-emergent herbicide (Bicep). All raw agronomic data and raw compositional data were analyzed by ANOVA in R using a linear model with location, block nested within location, and NIL genotype as fixed effects. We present p-values for genotype effects, and boxplots show residuals from a model including location and block effects.

*Fine-mapping and GWAS*

Two MITE-based indel markers were used to screen 1132 $BC_3$ and $BC_2F_2$ individuals. Subsequent genotyping-by-sequencing was performed in both putative recombinants and a GWAS panel of 1624 diverse sorghum accessions using the two-enzyme GBS protocol with PstI-HF and HinP1I enzymes (Poland et al., 2012), followed by alignment to v3 of the reference genome (www.phytozome.org) using bowtie2 (Langmead and Salzberg, 2012) and SNP calling using the TASSEL5 GBSv2 pipeline (Glaubitz et al., 2014). Imputation of biparental recombinants was performed using FSFHap (Swarts et al., 2014) in TASSEL, and imputation of the GWAS panel was performed using Beagle4 (Browning and Browning, 2007). Midrib color GWAS was performed using the mixed linear model implemented in GAPIT, using model selection to choose the optimal number of principal components, which was zero. A minor allele frequency cutoff of 5% resulted in 50,899 SNPs for testing, and only associations significant at a false discovery rate of less than 5% ($q < 0.05$) are reported. GBS data for this project have been deposited at the Illinois Data Bank (https://databank.illinois.edu/).

*Lignocellulosic compositional analysis*

Lyophilized tissue samples were ground to a fine powder using three 5mm metal balls in 2 mL plastic tubes (Retsch Ball mill, 2 times at 25 Hz for 2.5 minutes) and washed sequentially with 70 % ethanol, 1:1 (v:v) methanol : chloroform, and acetone. The alcohol-insoluble cell wall material was further destarched with alpha amylase (Sigma) and Pullulanase M2 (Megazyme) in 0.1 M Citrate buffer pH 5.0. The destarched material was aliquoted for different compositional assays. 1 mg of destarched cell wall was hydrolyzed in 2M Tri-fluoroacetic acid (TFA) heated to 121 ˚C for 90 minutes followed by a stream of dried air using nitrogen. Dried samples were re-

suspended in water, centrifuged, and the supernatant was collected for monosaccharides analysis and the pellet was dried for crystalline cellulose measurements. TFA-soluble samples were analyzed using High Performance Anion Exchange liquid Chromatography with Pulse Amperometric Detection (HPAEC-PAD) according to (de Souza et al., 2014). Neutral sugars were separated via a CarboPac PA20 column, while a CarboPac PA200 was used to separate uronic acids. Three distinct programs were used to resolve the sugars of interest. Samples were run at a flow rate of 0.4 mL / min and gradients consisted of i) 2mM NaOH for 20 min followed by a 5 minute 100mM flush and subsequent 5 minutes at 2mM (neutral sugar separation 1; excludes xylose and mannose); ii) 18 mM NaOH for 15 min followed by a 5 minute 100mM flush and subsequent 7 minutes at 18 mM (neutral sugar separation 2; excludes rhamnose and arabinose); iii) 0.1 M NaOH with a gradient of 50 – 200 mM sodium acetate from 0 to 10 minutes followed by a 2 minute 200 mM sodium acetate flush returning to 50 mM for 2.9 minutes (uronic acid separation). Crystalline cellulose was measured using the Updegraf method according to (Updegraff, 1969), and the released glucose was measured using the anthrone assay (Laurentin and Edwards, 2003). Lignin content and lignin composition were measured using the ultra-violet acetyl bromide lignin method and a thioacidolysis procedure respectively, according to (Foster et al., 2010). Saccharification yield of cell wall materials was determined after enzymatic treatment. In brief, 1 mg of destached cell wall was incubated with 0.5 µL Accellerase 1500 enzyme mix (Gencor) in 1 mL of 50 mM citrate buffer (pH 4.5), shaking at 250 rpm at 50 °C for 20 hours._Solubilized glucose and xylose were detected on a Bio-Rad HPX-87H, 300 x 7.8 mm column in an Shimadzu UFLC chromatography system. The elution profile encompassed 0.01 N sulfuric acid in 15 min at 0.6 ml/min and column temperature is 50 °C.

*Phylogenetic tree construction and sequence alignment*

*SbNAC074a* homologs from *Arabidopsis thaliana*, *Oryza sativa subsp. Japonica, Sorghum bicolor, Zea mays, Setaria italica,* and *Glycine max* were extracted from UNIPROT (http://www.uniprot.org). The full-length version of SbNAC074a, including the complete first exon, was used. Alignment was performed using MUSCLE in MEGA7 with default settings, and CLUSTALW2_Phylogeny was used with distance correction off, gaps excluded, and the UPGMA clustering method to create the tree. Visualization was performed using EVOLVIEW (http://www.evolgenius.info/evolview/#login).

*Gene expression analysis*

Total RNA was extracted from greenhouse-grown midrib tissue of the youngest fully-expanded leaf of *DD* and *dd* NILs at the four-leaf and six-leaf stages using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, USA), following the manufacturer's protocol. All RNA samples were digested with DNase I (New England Biolabs, USA), and rtPCR was performed using M-MuLV reverse transcriptase (New England Biolabs, USA). RNAseq libraries were prepared and sequenced at Roy J. Carver Biotechnology Center at the University of Illinois using single-end, 100 bp reads on a HiSeq2500 instrument. Reads were processed using HISAT2 and StringTie, and the stattest function in the R package "Ballgown" was used to test for differential gene expression between *DD* and *dd* NILs using 3 biological replicates at each developmental stage, controlling the false discovery rate at 0.05. The R package "WGCNA" was used to construct co-expression modules using the arguments power = 8, minModuleSize = 20 , and mergeCutHeight = 0.05. RNAseq reads for this project have been deposited at the Illinois Data Bank (https://databank.illinois.edu/).

## RESULTS

*Creation and characterization of nearly-isogenic lines for the D locus*

A white midrib allele *(D)* was introduced into the green midrib Tx623 background *(dd)* through four generations of backcrossing, and homozygous *DD* and *dd* seed stocks were derived from a single $BC_4$ plant by selfing (Figure 1). NILs appear identical until the sixth-leaf stage, when a narrow band of pithy white tissue first appears in *DD* midribs, becoming wider and more distinct in successive leaves (Figure 2A). At anthesis, *DD* stalks are visibly drier and pithier than *dd* stalks (Figure 2B), and stem sections show that *DD* stalks have significantly more vascular bundles ($p < 0.001$, Figure 2C-D).

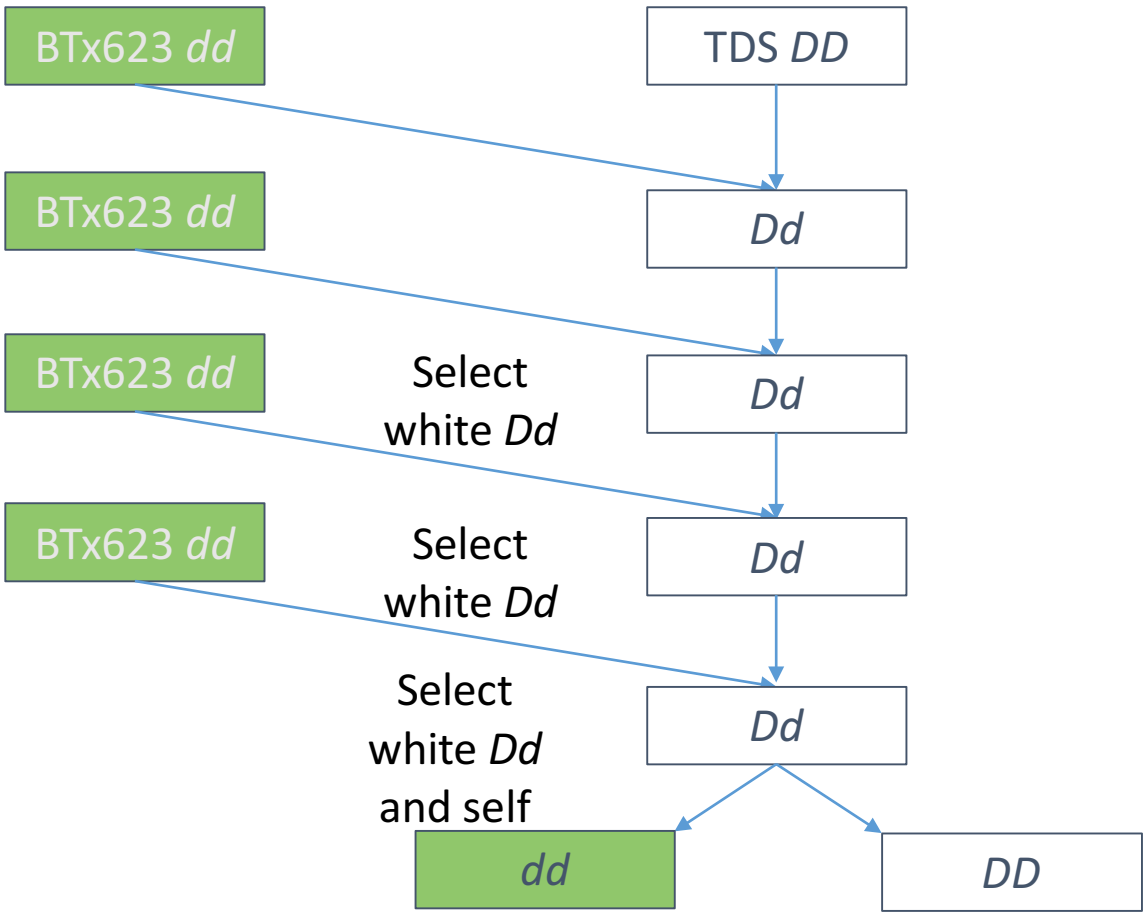


**Figure 1.** the process of creation of the NILs by four generations of backcrossing and homozygous *DD* and *dd* seed stocks were derived from a single $BC_4$ plant by selfing.

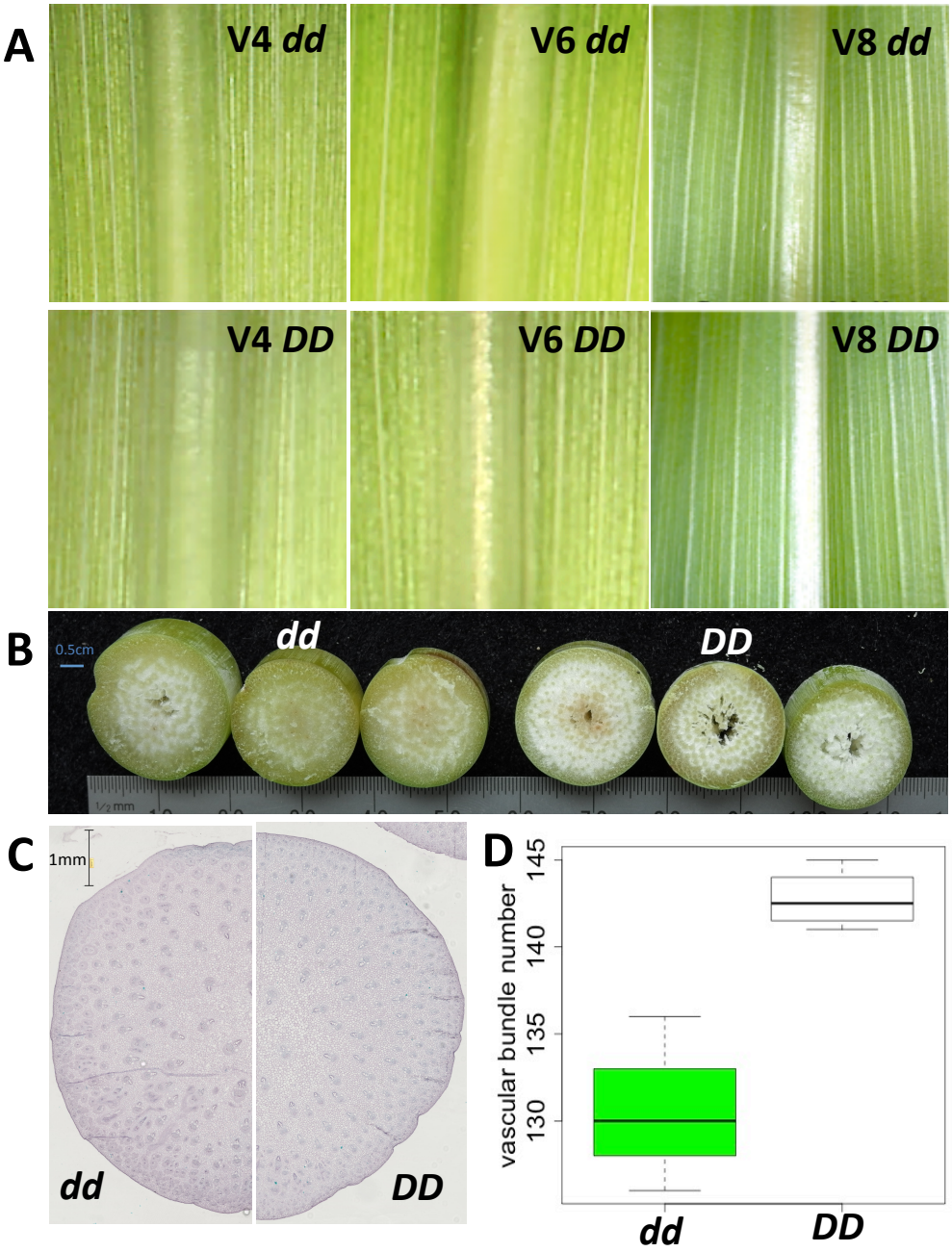


**Figure 2.** Phenotypic differences between *DD* and *dd* NILs. **A**: Midribs of the youngest fully-expanded leaf in *dd* and *DD* NILs at V4 , V6, and V8. **B**: Cross-sections through the 2$^{nd}$ stalk internode above ground level at 10 weeks after anthesis, with 0.5cm scale bar. **C**: cross-section

of young expanding internode 6 weeks after planting, with 1mm scale bar. **D**: Number of vascular bundles in 5-mm internodes of *dd* and *DD* NILs at the V8 stage.

*GWAS and fine-mapping*

Scoring midrib color as a binary trait (green/white) in a large panel (n=1624) of sorghum inbreds reveals a single very strong association at ~ 51 Mb on chromosome 6 (Fig 3A). Polymorphic markers flanking this region were used to screen 1132 $BC_3$ and $BC_2F_2$ individuals. Putative recombinant individuals were subjected to genotyping-by-sequencing, confirming 17 recombinants and defining a ~36 kb interval that co-segregates perfectly with the phenotype (Fig 3B). This interval contains four predicted genes, of which two are expressed in midrib tissue at either the fourth- or sixth-leaf stage: a NAC transcription factor (Sobic.006G147400) and a threonine aldolase (Sobic.006G147600). The two most significant hits in the GWAS analysis are the two closest flanking SNPs to the NAC gene. Moreover, the NAC gene is the only gene in the interval with a different annotated exon-intron structure compared to its closest homologs in other cereals, which is relevant because the sorghum reference genome is derived from Tx623, a recessive *dd* mutant. Reverse-transcriptase PCR and cDNA sequencing confirms that the annotated first intron in Sobic.006G147400 does not exist, and that a T/C SNP in this region produces a stop codon in the *d* allele from Tx623 but not the contrasting D allele (Figure 3C). The premature stop codon in Tx623 eliminates most of the NAC domain.

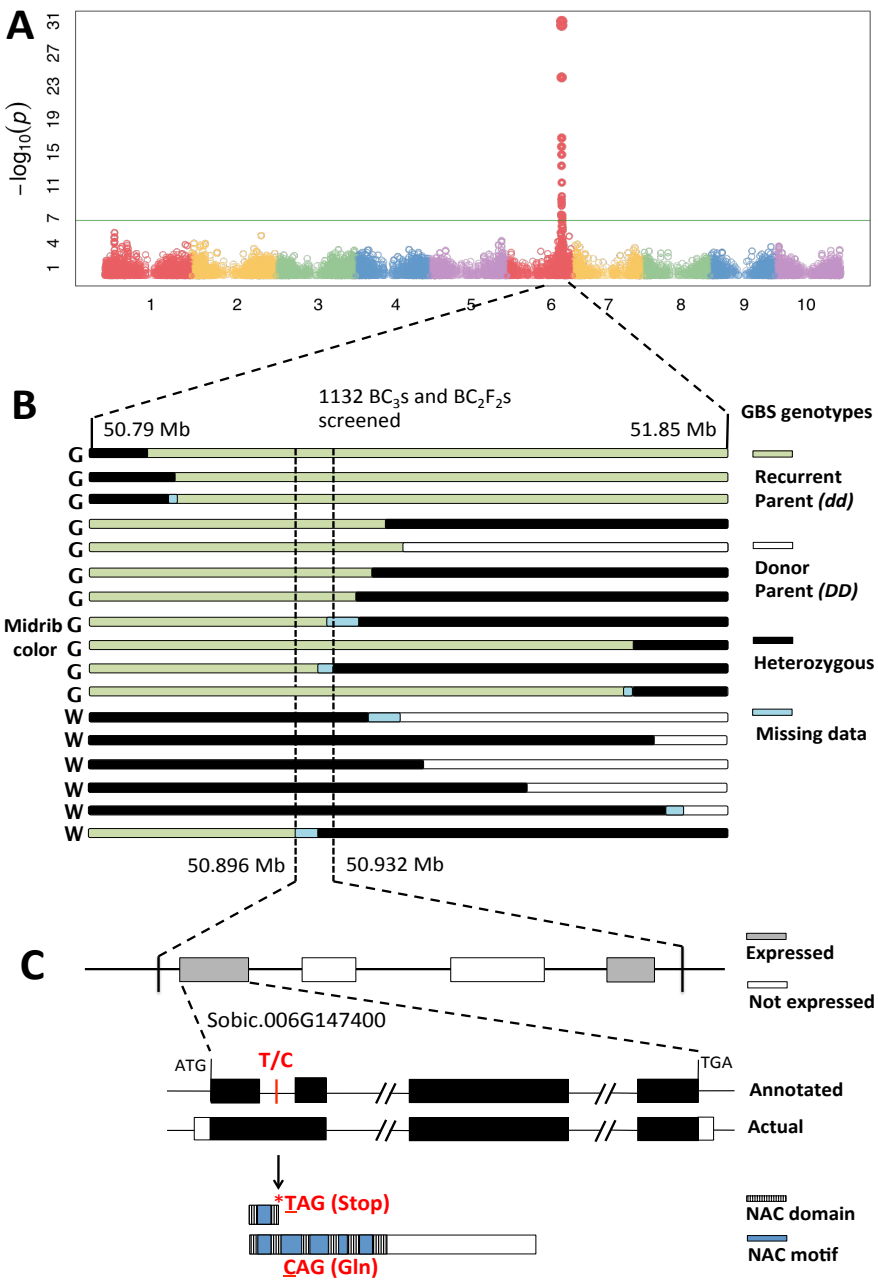


**Figure 3.** GWAS and fine-mapping of the sorghum *D* locus. **A**: Midrib color GWAS in 1624 sorghum inbreds, with a major peak at ~51 Mb on chr6. **B**: Screening 1132 individuals with flanking markers yielded 17 recombinants, which were GBSed to define a 36 kb interval. **C:** Two of four genes in the 36 kb interval are expressed, one of which is a NAC transcription factor (Sobic.006G147400) with a stop codon in the first exon disrupting the NAC domain. This gene

was annotated incorrectly in the sorghum reference genome (Tx623), which has the stop codon and a recessive *d* allele.

*Phylogeny and identification of homologs*

The full-length NAC protein, derived from an allele of Sobic.006G147600 lacking the premature stop codon, was used to search for homologous proteins in four grasses (*Oryza sativa subsp. japonica, Sorghum bicolor, Zea mays, Setaria italica)* and two dicots (*Arabidopsis thaliana* and Glycine max), which were aligned using MUSCLE in MEGA7 and used to construct a neighbor-joining tree using UPGMA clustering (Figure 4). Syntenic regions on *Oryza* chromosome 4 and *Setaria* chromosome 7 each contain single orthologous copies of our sorghum NAC candidate, and co-syntenic regions on *Zea* chromosomes 2 and 10 each contain intact homeologues derived from segmental duplication. These proteins are part of the NAC1 sub-clade (Peng et al., 2015b), and contain five conserved motifs in the N-terminal NAC domains whereas the C-termini are highly variable (Figure A1). The closest *Arabidopsis* homolog is NAC074 (At4G28530), and we hereafter refer to the three co-orthologous grass clades as NAC074a, NAC074b, and NAC074c, and to Sobic.006G147400 as *SbNAC074a*. In both NAC074a and NAC074c clade there are two copies of maize genes, while there is only one copy of maize gene in NAC074b clade. This implies a is lost after the genome duplication of maize.

*dd NILs show increases in stalk dry matter, soluble sugar yield, and grain yield*

To confirm the previously-reported association between the D locus and sugar yield, we grew NILs in six replications of paired rows, extracted total sugar from 1 m of row 30 days after

anthesis, and found that *dd* NILs yield nearly twice the sugar of *DD* NILs (Figure 5A, $p < 0.001$) despite being identical in plant height and flowering time. Sugar yield per meter of row (g m-1) is a function of juice volume (ml m-1) and the concentration of simple sugars in the juice (Brix; g ml-1), and we observe that increased yield in *dd* NILs is driven by increased juice volume, which in turn is associated with higher vegetative weight (Figure 5B-D). Given the dramatic effect of the *D* locus on sugar yield, we next conducted a time series experiment to quantify its effects on the rate and extent of grain filling. Paired, two-row plots of NIL were grown in four replicates at each of two locations, and dry matter and moisture of stalk and grain were monitored at two-week intervals beginning at anthesis. Overall, this grain filling period is characterized by movement of dry matter from the stalk to the grain, and by decreases in grain moisture relative to stalk moisture (Figure 6). Stalk dry matter is significantly higher in *dd* NILs at all stages, most notably at 4 weeks after anthesis when it is >58% higher ($p = 0.004$; Figure 5A). *dd* NILs also have higher grain moisture at 4 weeks after anthesis ($p = 0.042$; Figure 5D), and higher grain yield at 6 weeks after anthesis ($p = 0.025$; Figure 6B).

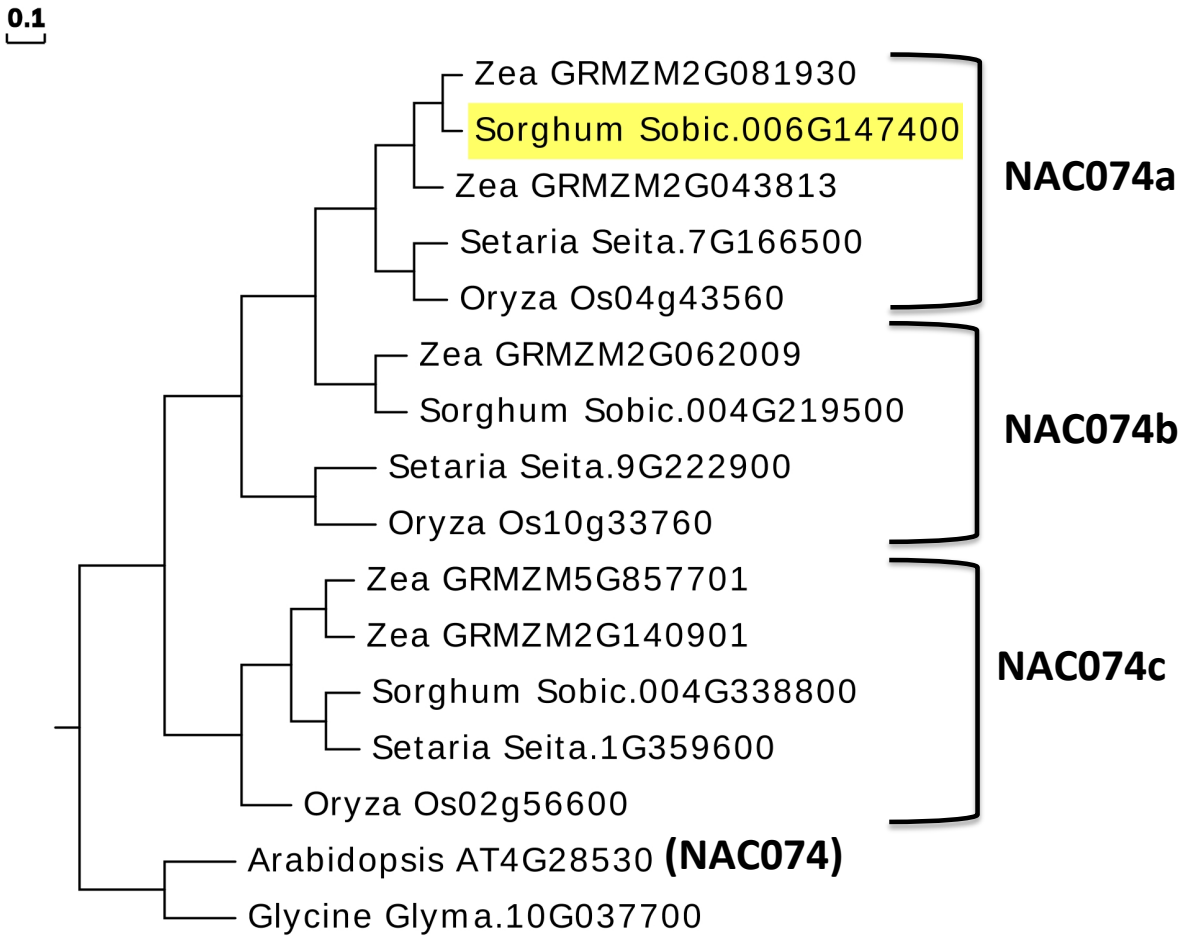


**Figure 4.** UPGMA tree of NAC074 genes in grasses. *SbNAC074a* is highlighted.

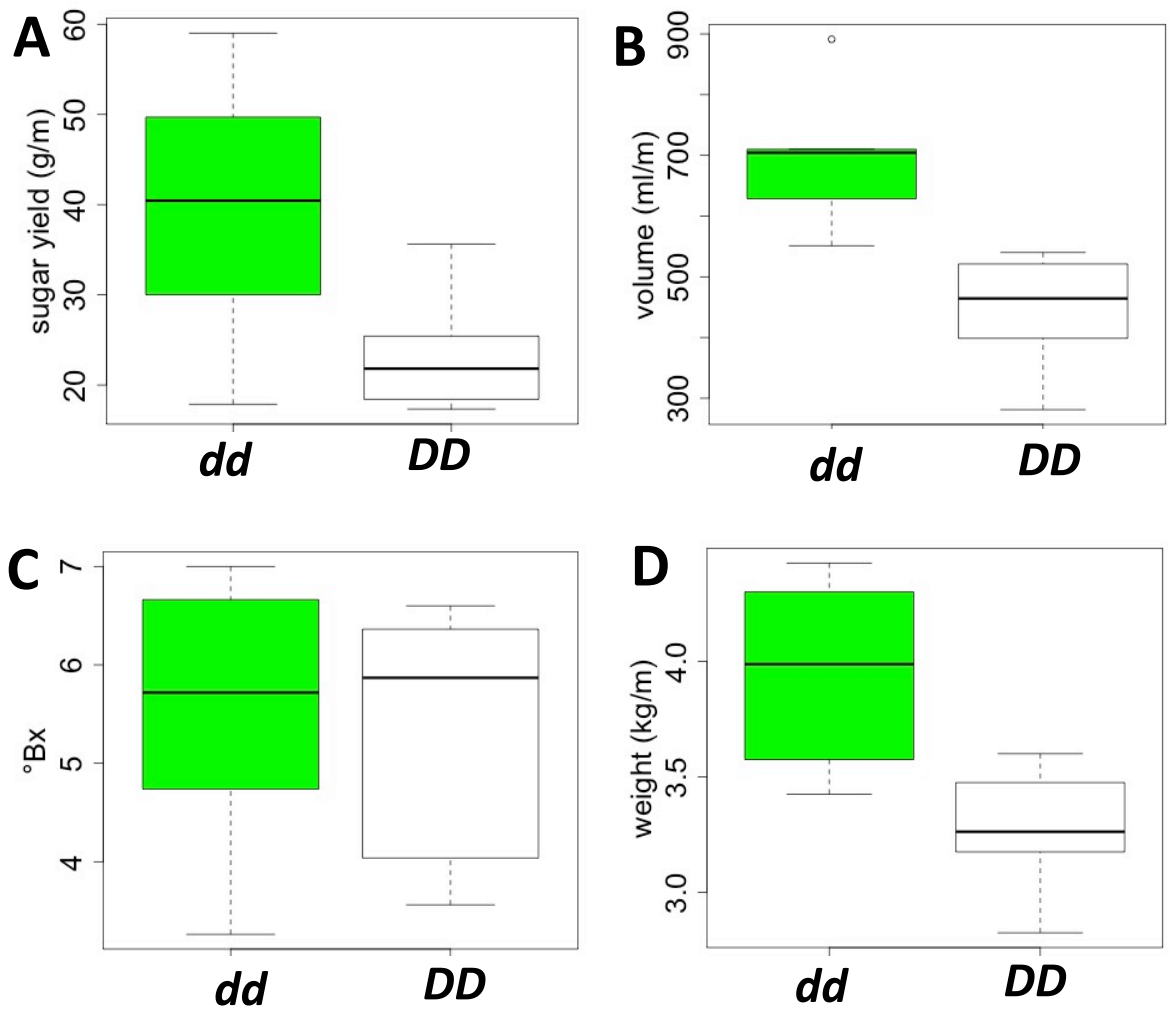


**Figure 5.** Sugar accumulation in *DD* and *dd* NILs. **A**: Sugar yield (g m-1); **B**: Juice volume (ml m-1); **C**: Brix (g ml-1); **D**: Vegetative Wet Weight (kg m-1). Sugar yield is calculated by multiplying juice volume and brix.

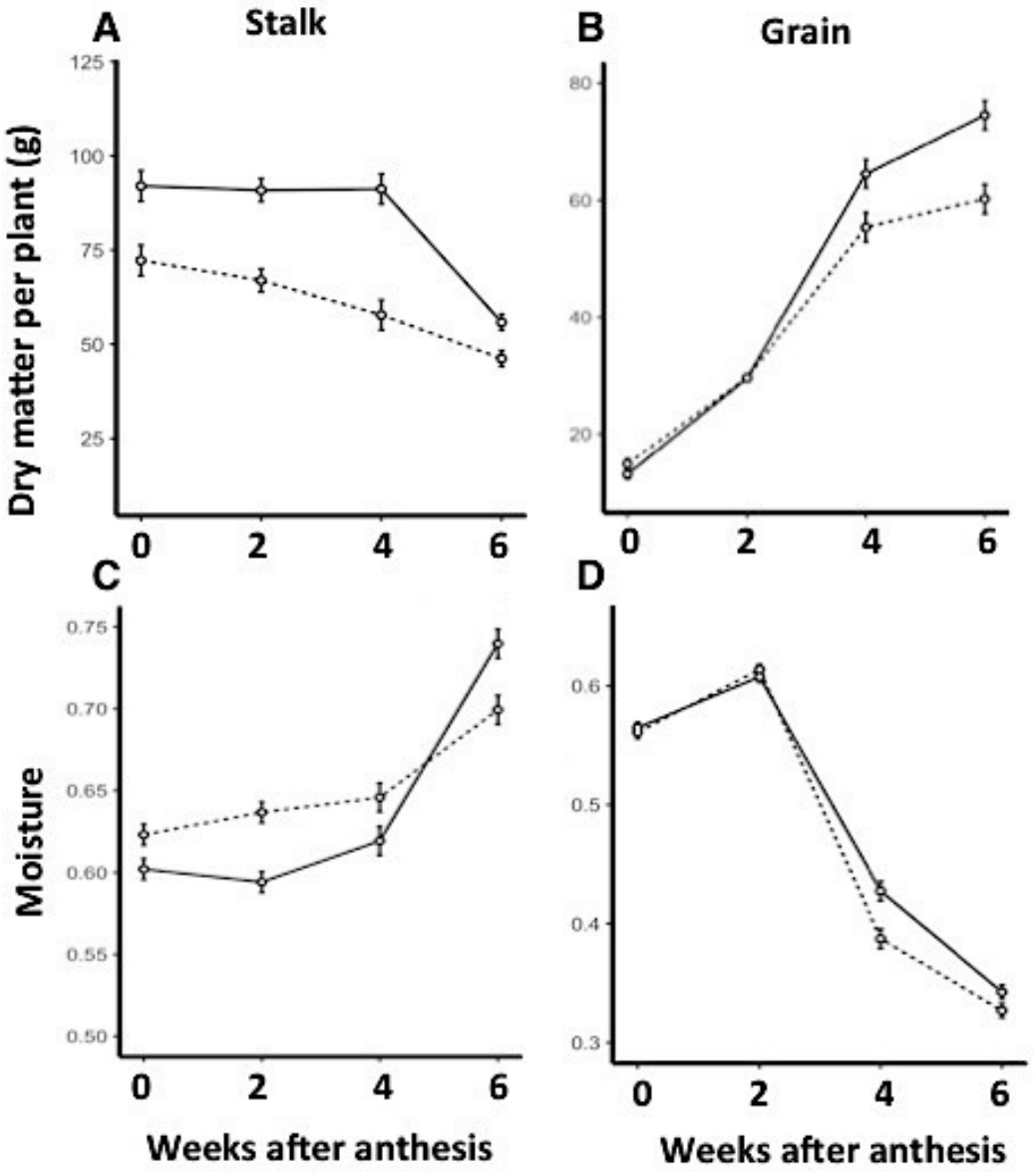


**Figure 6.** Grain filling of *DD* NILs (dashed lines) and *dd* NILs (solid lines) under field conditions. Changes in dry matter (A, B) and moisture (C, D) were monitored at 2-week intervals in stalk (A, C) and grain (B, D) from 0-6 weeks after anthesis.

*Compositional differences drive increased biomass digestibility of dd NILs*

NIL stalk tissue was subjected to detailed lignocellulosic compositional analysis at two growth stages: developing internodes from a 1 cm plug of tissue immediately below the shoot apical meristem at 6 weeks after planting (stage 1), and the third internode below the inflorescence at 9 weeks after planting, which coincided with the boot stage shortly before flowering (stage 2). While stage 1 tissues show relatively few compositional differences between NILs, stage 2 *dd* NILs show reduced lignin (p=0.007; Figure 7A), increased glucose release following enzymatic saccharification (p=0.049; Figure 7B), and a marginally significant decrease in crystalline cellulose (p=0.055; Figure 7C). *dd* NILs also show a lower ratio of syringyl to guaiacyl lignin monomers at stage 2 (p=0.029), and differences in acid-hydrolyzed lignocellulosic monosaccharides at both stages (Figure A2).

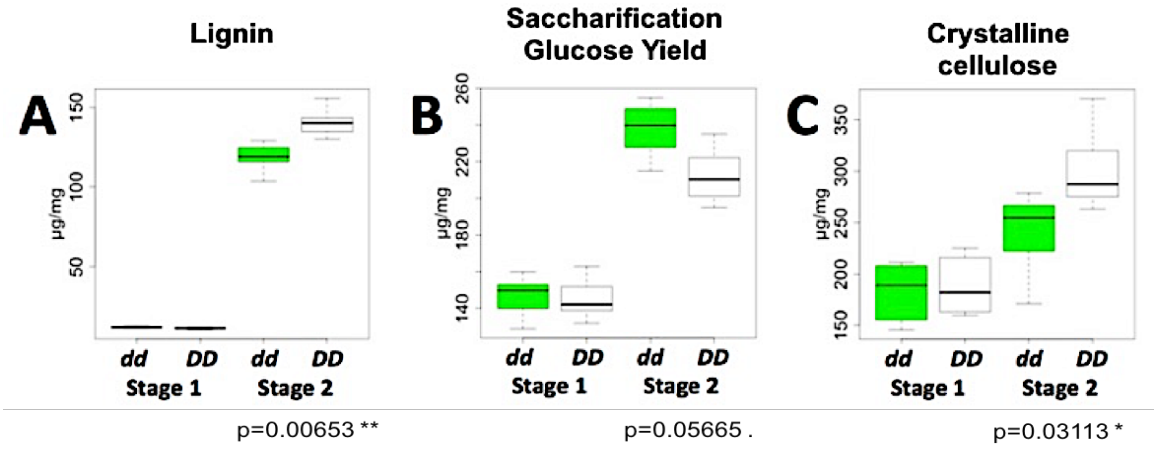


**Figure 7.** Compositional analysis of developing internodes in *DD* and *dd* NILs. **A**: Lignin (p=0.0065); **B**: Saccharification glucose yield (p=0.0311); **C**: Crystalline cellulose (p=0.0565). Tissues sampled included developing internodes at 6 weeks after planting (Stage 1) and the third internode below the inflorescence at 9 weeks after planting (Stage 2).

*Transcriptome profiling suggests SbNAC074a represses a miniature zinc finger (MIF) gene*

Transcriptome profiling was performed on mRNA from NIL midrib tissue at the four-leaf and six–leaf stages, just before and during the first appearance of phenotypic differences between the NILs (Figure 1A). Three independent biological replicates were obtained from pooled tissue from three separate greenhouse plantings, for a total of 12 samples across the two NILs and two stages. Co-expression analysis using WGCNA (Langfelder and Horvath, 2008) shows that *SbNAC074a* shares a regulatory module with 263 other annotated genes, including a group of five NAC transcription factors (Figure 8) that includes *SbNAC074c* but not *SbNAC074b* as well as two NACs associated with xylem development in Arabidopsis, *Xylem NAC Domain 1 (XND1)* and *NAC075* (Endo et al., 2015; Zhao et al., 2007). However, using the HiSAT-Stringtie-Ballgown pipeline (Pertea et al., 2016), only a single gene (Sobic.008G020700) is differentially expressed between DD and *dd* NILs ($q<0.01$), showing no expression in *DD* NILs and mean expression of 21 and 35 FPKM in *dd* NIL midribs at the four-leaf and six-leaf stages, respectively. Sobic.008G020700 is annotated as a MIF gene, which are seed plant-specific, truncated versions of ZF-HD transcription factors (Hu and Ma, 2006) that dimerize with ZF-HDs and suppress their transcriptional activation activity (Hong et al., 2011). Intriguingly, the only significant GWAS hit for midrib color other than our candidate NAC gene falls near a ZF-HD transcription factor on chromosome.

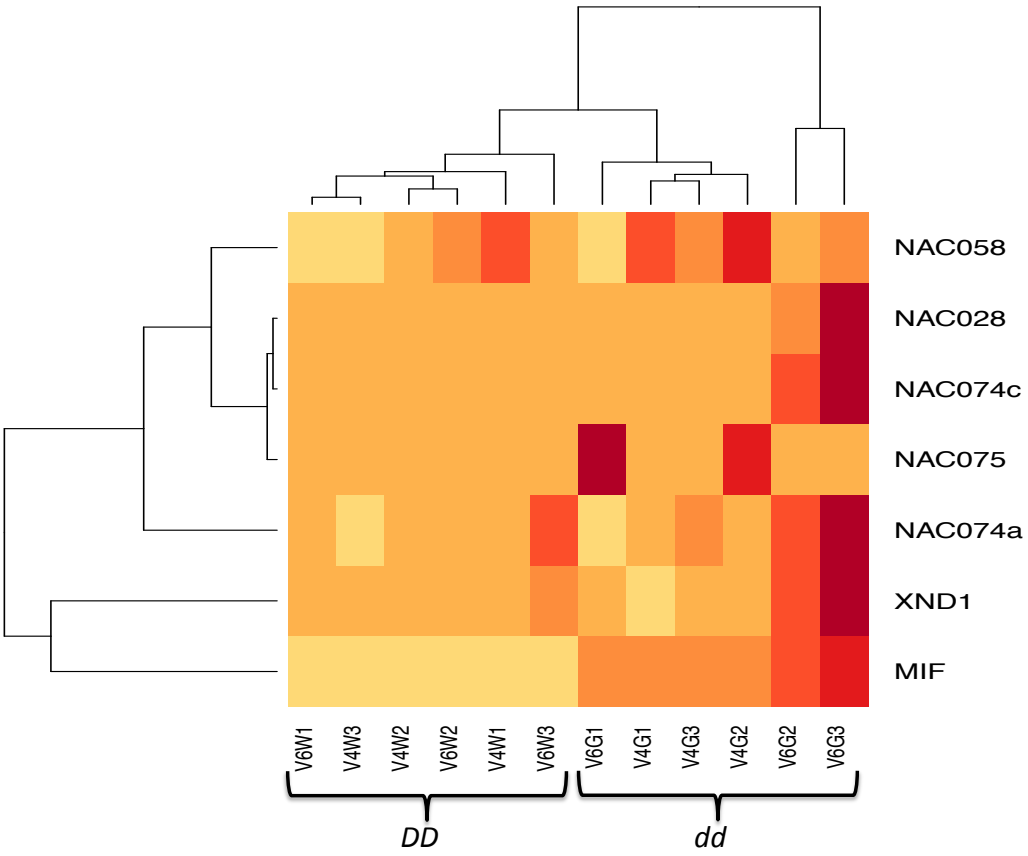


**Figure 8.** A. Co-expression of SbNAC074a, the differentially-expressed MIF gene, and 5 other NAC transcription factors in V4 and V6 midribs of *DD* and *dd* NILs. Light and dark colors indicate low and high gene expression respectively. Dendrograms reflect Euclidean distance and determine row and column order.

## DISCUSSION

In this study, we fine-map the sorghum *D* locus to a four-gene interval that includes a NAC transcription factor with a premature stop codon in the *dd* sorghum reference genome. In addition, we note that the two most significant SNPs in the GWAS analysis are the two closest flanking markers to this NAC gene in our dataset. While we have not formally demonstrated that this NAC gene underlies the *D* locus in sorghum, we present anatomical, compositional, transcriptional, and agronomic data from *dd* and *DD* NILs that is consistent with perturbation of NAC gene function. We have named this gene *SbNAC074a*, after its closest homolog in Arabidopsis.

Previous study of genes orthologous to and co-expressed with *SbNAC074a* suggest a role in xylem development. *Arabidopsis* NAC074 (At4G28530) is upregulated in xylem relative to phloem-cambium and non-vascular tissues (Zhao, 2005), and is downregulated in plants overexpressing ANAC012/SND1, a key regulator of xylary fiber development (Ko et al., 2007). The rice ortholog of *SbNAC074a* (Os04g43560) is upregulated in panicle and root under drought stress in drought-tolerant, but not drought-susceptible, NIL backgrounds (Nuruzzaman et al., 2012b). Here we show that *SbNAC074a* is co-expressed with other NAC transcription factors involved in xylem development, including homologs of Arabidopsis *XND1* and *NAC075*. *xnd1* mutants display a mild dwarfing phenotype associated with a reduction in tracheary element length, whereas overexpression of XND1 results in reduced formation of xylem vessels, expansion of the phloem, and increased starch storage in amyloplasts (Zhao et al., 2007).The authors suggest that XND1 may promote vessel elongation by repressing their terminal differentiation. Overexpression of NAC075 results in ectopic formation of xylem vessel elements

(Endo et al., 2015) and rescues the pendent stem phenotype of *nst1-nst3* double mutants, which results from complete loss of secondary cell wall deposition in xylem fibers.

Lignocellulosic compositional and agronomic data are consistent with a role for *SbNAC074a* as a positive regulator of xylem development. *dd* NIL internodes have significantly reduced lignin. Lignin content has been positively correlated with xylem development and inversely correlated with lignocellulosic saccharification yields (Chen and Dixon, 2007), consistent with the observed increase in glucose yield following enzymatic saccharification in *dd* NILs. The increased digestibility of *dd* NILs comes with no obvious agronomic penalty in the dwarf grain sorghum background of Tx623, though we observe that green midrib accessions have higher stalk lodging than white midrib accessions in a biomass sorghum panel (Figure A3). Strikingly, yields of soluble sugar, grain, and vegetative biomass are all significantly increased in *dd* NILs under well-watered field conditions. Loss of xylem-identity genes can lead to increased phloem proliferation (Zhao et al., 2007, p. 1), and vice versa (Bonke et al., 2003). We propose that loss of *SbNAC074a* function in *dd* NILs alters phloem development, increasing stalk carbon reserves to sustain a longer grain filling period.

Transcriptome profiling suggests that *SbNAC074a* represses a MIF gene, implicating MIF-targeted ZF-HDs in the control of vascular architecture in grasses. In *Arabidopsis*, the promoter of the drought-responsive ERD1 gene contains both NAC and ZF-HD binding sites, and both NAC and ZF-HD proteins are required activate expression of ERD1 (Tran et al., 2007). While the sorghum *D* locus on chromosome 6 is by far the most significant GWAS association for midrib color, the second most significant association lies close a ZF-HD gene on chromosome 1. We propose that *SbNAC074a* represses Sobic.008G020700, rescuing ZF-HD TFs from dimerization with a repressive MIF protein, and that *SbNAC074a* and/or other NACs

including XND1 act in concert with ZF-HD proteins for activation of downstream genes promoting xylary tissue identity.

# CHAPTER TWO: DIFFERENTIAL GENE EXPRESSION IN MIDRIB AND STALK TISSUE OF NILs FOR THE SORGHUM D LOCUS


## SUMMARY

A differential gene expression (DGE) study on sorghum stalk and midrib samples was carried out in NILs segregating for the Dry Midrib (D) locus. There were three sets of analyses. The first set included V4 and V6 midrib samples from greenhouse-grown plants previously described in chapter one. The second set included 6- and 9-week-old field-grown stalk samples. The third set was the combination of set 1 and set 2. DGE analyses were performed using R package *edgeR* with *glmFit* analysis and *blast2go*. *glmFit* provides more lenient FDR control compared with *ballgown,* which was used in chapter one. Using *glmFit*, 5 genes showed DGE in all three sets but only Sobic.006G145600, a glycosyl hydrolase gene, had annotation. Sobic.006G145600 is physically linked to the *D* locus at a distance of ~160 kb from the NAC074a gene (Sobic.006G147400). The glycosyl hydrolase lies at ~50.74 Mb on chromosome 6 whereas NAC074a lies at ~50.90 Mb, and we cannot exclude the possibility that the NILs used in this experiment are segregating for Sobic.006G145600 as well as Sobic.006G147400. GO analysis of DGE ($q<0.1$) showed glycosyl hydrolase activity was significantly different between *dd* and *DD* in all sets. At this point, we cannot fully distinguish between *cis*- and *trans*- effects on glycosyl hydrolase expression in sorghum NILs for the D locus. Pairwise analysis also suggested the MIF gene, Sobic.008G020700, was upregulated in *dd* midribs.

# METHODS

*RNA extraction and isolation*

Methods of midrib RNA sample preparation are described in chapter two. Stalk RNAs were extracted from field-grown stalks using the same tissues and sampling times described for cell wall compositional analysis in chapter 3. Briefly, sampling occurred at 6 weeks (S1) after sowing and 9 weeks (S2) after sowing. At S1 stage, a 1 cm plug of tissue including young expanding internodes and the apical meristem were collected. At S2 stage, 1 cm sections of peduncle were collected. Paired 2-row plots of NILs were planted in four replicates at each of two locations (Urbana and Savoy, IL), and samples were pooled from 2 individual representative plants per row for each time point. Total RNA was extracted using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, USA), following the manufacturer's protocol. All RNA samples were digested with DNase I (New England Biolabs, USA).

*Reads Alignment and Analysis*

RNAseq libraries were prepared and sequenced at Roy J. Carver Biotechnology Center at the University of Illinois using single-end, 100 bp reads on HiSeq2500 and HiSeq4000 instruments. Reads of 24 samples were aligned and assembled using the HiSATII and StringTie pipeline. EdgeR was used to perform DGE for midrib, stalk and total samples individually using *glmFit* analysis. Pairwise DGE is performed using *edgeR* R code in set 1 and set 2. Gene ontology (GO) study was performed on AGRIgo using the *Sorghum bicolorV2.1* annotation (Du et al., 2010; Langfelder and Horvath, 2008; Robinson et al., 2010).

# RESULT

*Differential Gene Expression*

Genotype and sampling time were the two factors in the general linear model to analyze set 1 and set 2 and set Genotype and tissue were the two factors in the general linear model to analyze set 3. We initially used *glmQLTest* to compare the DGE between genotypes. There was only one DGE in the midrib analysis (set 1), Sobic.008G020700, which confirmed the DGE result using ballgown in chapter two. There were three and five DGEs in the stalk analysis (set 2) and combined analysis (set 3) respectively.

We next used *glmFit* to perform a less stringent DGE analysis. Genotype and sampling time were still used as factors for set 1 and set 2 and genotype and tissue were still used as factors for set 3 in the general linear model but *glmFit* was used to compare the DGE between genotype. There were 105, 50 and 60 DGEs in midrib, stalk and total samples individually. Five genes showed DGE in all three analyses, but only one gene, Sobic.006G145600, had gene annotation. This gene was annotated as glycosyl hydrolase. It was 3.3, 5.8 and 4.5 fold upregulated in the *dd* genotype in midribs, stalks, and the combined analysis respectively.

Pairwise DGE analysis resulted in 539, 224, 474 and 314 genes upregulated in *DD* in V4 midrib, V6 midrib, S1 stalk and S2 stalk respectively, and 373, 449, 300 and 297 genes upregulated in *dd* in V4 midrib, V6 midrib, S1 stalk and S2 stalk. We summarized the common DGE of *DD* and *dd* in V4 and V6 and S1 and S2 independently. Sobic.008G020700 was upregulated in *dd* midribs. An allyl alcohol dehydrogenase, Sobic.003G110200, was upregulated in all *dd* tissues. No annotated genes were commonly upregulated in all *DD* tissues.

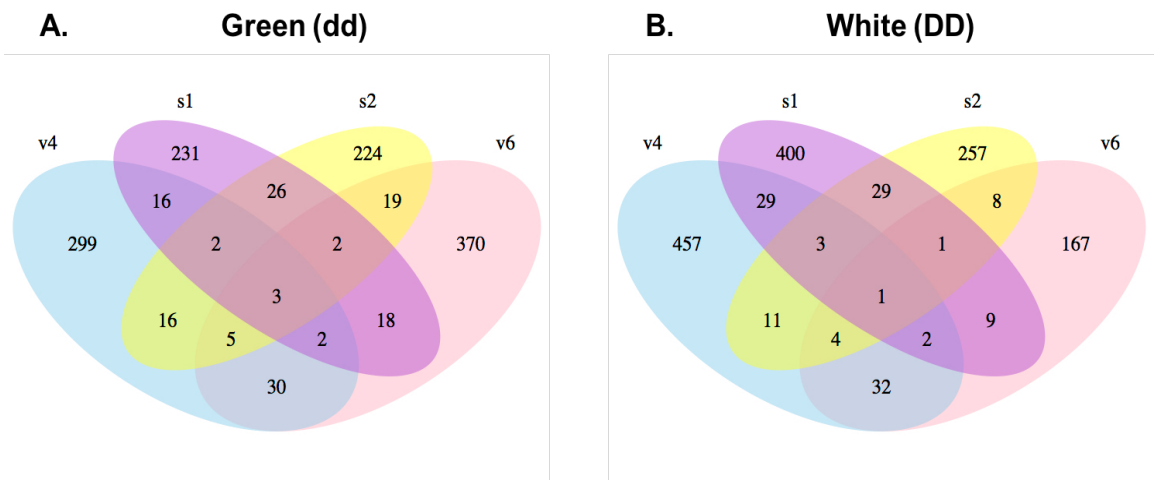


**Figure 9. Venn diagram of upregulated genes in A: *dd* and B: *DD* NILs.**

*Gene Ontology Study*

Output from *glmFit* was used to perform a gene ontology (GO) study. There was no significant GO term common to all three sets with an FDR cutoff of 0.05. When we relaxed the threshold to FDR<0.2, both midrib and midrib + stalk sets had two identical significant GO terms and the stalk set had no significant GO term. The two significant GO term were GO:0004553 and GO:0016798. GO:0004553 and GO:0016798 both involve hydrolase activities. GO:0004553 hydrolyzes o-glycosyl compounds and GO:0016798 acts on glycosyl bonds.

## DISCUSSION

In this chapter, we identify Sobic.006G145600 as a glycosyl hydrolase gene that shows DGE between green (dd) and white (DD) NILs in midrib, stalk, and midrib + stalk analyses. Its Arabidopsis ortholog, AT1G061810, is $\beta$-GLUCOSIDASE45 (BGLU45) gene that belongs to Glycosyl Hydrolase Family 1 (GH1). BGLU45 is a $\beta$-glucosidase that is specific for coniferin, syringin, and $\rho$-coumaryl alcohol 4-O- $\beta$ -D-glucoside and is found in interfascicular fibers. *In vitro*, BGLU45 is expressed in lignifying tissues and organs. Arabidopsis mutants with BGLU45 knocked out show decreases in syringin. Although *in silico*, co-expression analysis showed BGLU45 may be involve in lignin synthesis, Arabidopsis *bglu45/bglu45* mutants have no significant changes in stem lignin content or S, G and H lignin level. The lack of significant effect on lignin content in BGLU45 mutants may be explained by the fact that BGLU45 accounts for less than 8% of glucosidase activity in Arabidopsis mature stem (Peng et al., 2015a; Xiao et al., 2016). GO analysis also indicated glycosyl hydrolase genes are differentially expressed between the NILs. Previous evidence from a maize mapping study suggests that an O-glycosyl hydrolase may be associated with stalk strength (Peiffer et al., 2013). However, Sobic.006G145600 (glycosyl hydrolase) is physically so close to Sobic.006G147400 (NAC74) that we cannot separate their effects.

# CONCLUSION

The *D* locus associated with sorghum green/white midrib color was validated by GWAS using 1624 inbreds and the interval was fine mapped using flanking markers and GBS on 1132 $BC_3$ and $BC_2F_2$ individuals. Four genes were in the fine mapped region and only two of these genes were expressed in midrib tissue. Compared with their orthologs from other grass species, we found that only a NAC74 gene, which was in the fine mapped interval, had a different gene model than its orthologs from other grass species. By sequencing this gene in *dd* and *DD*, we found a premature stop codon that may result in the midrib color polymorphism in *dd* and *DD*. *dd* NILs had fewer vascular bundles, higher cell wall digestibility, more biomass accumulation and extended grain filling stage compared to *DD* NILs. Expression analysis of the *dd* and *DD* NILs showed increased expression of a MIF gene in *dd* midrib tissue. MIF proteins form heterodimers with ZF-HD proteins and consequently prevent ZF-HD proteins from functioning as homodimers. DGE studies indicated that increased glycosyl hydrolase expression in *dd* may be partially responsible for the phenotypic variation between the NILs in both stalk and midrib samples.

# REFERENCES


Bonke, M., Thitamadee, S., Mahonen, A.P., Hauser, M.T., Helariutta, Y., 2003. APL regulates vascular tissue identity in Arabidopsis. Nature 426, 181–186. doi:10.1038/nature02100

Browning, S.R., Browning, B.L., 2007. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. Am. J. Hum. Genet. 81, 1084–1097. doi:10.1086/521987

Burks, P.S., Kaiser, C.M., Hawkins, E.M., Brown, P.J., 2015. Genomewide Association for Sugar Yield in Sweet Sorghum. Crop Sci. 55, 2138–2148. doi:10.2135/cropsci2015.01.0057

Chen, F., Dixon, R.A., 2007. Lignin modification improves fermentable sugar yields for biofuel production. Nat. Biotechnol. 25, 759–761. doi:10.1038/nbt1316

Cosgrove, D.J., 2005. Growth of the plant cell wall. Nat. Rev. Mol. Cell Biol. 6, 850–861. doi:10.1038/nrm1746

de Souza, A., Hull, P.A., Gille, S., Pauly, M., 2014. Identification and functional characterization of the distinct plant pectin esterases PAE8 and PAE9 and their deletion mutants. Planta 240, 1123–1138. doi:10.1007/s00425-014-2139-6

Du, Z., Zhou, X., Ling, Y., Zhang, Z., Su, Z., 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. 38, W64–W70. doi:10.1093/nar/gkq310

Endo, H., Yamaguchi, M., Tamura, T., Nakano, Y., Nishikubo, N., Yoneda, A., Kato, K., Kubo, M., Kajita, S., Katayama, Y., Ohtani, M., Demura, T., 2015. Multiple classes of transcription factors regulate the expression of VASCULAR-RELATED NAC-DOMAIN7, a master switch of xylem vessel differentiation. Plant Cell Physiol. 56, 242–254. doi:10.1093/pcp/pcu134

Foster, C.E., Martin, T.M., Pauly, M., 2010. Comprehensive Compositional Analysis of Plant Cell Walls (Lignocellulosic biomass) Part I: Lignin. J. Vis. Exp. doi:10.3791/1745

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., Buckler, E.S., 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS ONE 9, e90346. doi:10.1371/journal.pone.0090346

Godin, B., Nagle, N., Sattler, S., Agneessens, R., Delcarte, J., Wolfrum, E., 2016. Improved sugar yields from biomass sorghum feedstocks: comparing low-lignin mutants and pretreatment chemistries. Biotechnol. Biofuels 9, 1–11. doi:10.1186/s13068-016-0667-y

Grain: World Markets and Trade | USDA Foreign Agricultural Service [WWW Document], n.d. URL https://www.fas.usda.gov/data/grain-world-markets-and-trade (accessed 3.20.17).

Hong, S.-Y., Kim, O.-K., Kim, S.-G., Yang, M.-S., Park, C.-M., 2011. Nuclear Import and DNA Binding of the ZHD5 Transcription Factor Is Modulated by a Competitive Peptide Inhibitor in Arabidopsis. J. Biol. Chem. 286, 1659–1668. doi:10.1074/jbc.M110.167692

Hu, W., Ma, H., 2006. Characterization of a novel putative zinc finger gene *MIF1* : involvement in multiple hormonal regulation of Arabidopsis development. Plant J. 45, 399–422. doi:10.1111/j.1365-313X.2005.02626.x

Jihong Wu, Yang Zheng, Baoguo Sun, Xiaotao Sun, Jiyuan Sun, Fuping Zheng, Mingquan Huang, 2015. The Occurrence of Propyl Lactate in Chinese Baijius (Chinese Liquors) Detected by Direct Injection Coupled with Gas Chromatography-Mass Spectrometry. Molecules 20, 19002–19013. doi:10.3390/molecules201019002

Kazuo Nakashima, Hironori Takasaki, Junya Mizoi, Kazuo Shinozaki, Kazuko Yamaguchi-Shinozaki, 2012. NAC transcription factors in plant abiotic stress responses. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1819, 97–103. doi:10.1016/j.bbagrm.2011.10.005

Ko, J.-H., Yang, S.H., Park, A.H., Lerouxel, O., Han, K.-H., 2007. ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in Arabidopsis thaliana. Plant J. 50, 1035–1048. doi:10.1111/j.1365-313X.2007.03109.x

Kumar, M., Campbell, L., Turner, S., 2016. Secondary cell walls: biosynthesis and manipulation. J. Exp. Bot. 67, 515–531. doi:10.1093/jxb/erv533
Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. doi:10.1186/1471-2105-9-559
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. doi:10.1038/nmeth.1923
Laurentin, A., Edwards, C.A., 2003. A microtiter modification of the anthrone-sulfuric acid colorimetric assay for glucose-based carbohydrates. Anal. Biochem. 315, 143–145.
Ma, Y., Sawhney, V.K., Steeves, T.A., 1993. Staining of paraffin-embedded plant material in safranin and fast green without prior removal of the paraffin. Can. J. Bot. 71, 996–999. doi:10.1139/b93-114
Mallory, A.C., Dugas, D.V., Bartel, D.P., Bartel, B., 2004. MicroRNA Regulation of NAC-Domain Targets Is Required for Proper Formation and Separation of Adjacent Embryonic, Vegetative, and Floral Organs. Curr. Biol. 14, 1035–1046. doi:10.1016/j.cub.2004.06.022
More, A.W., Kalpande, H.V., Ingole, D.G., Nirde, A.V., 2016. Heterosis studies for grain yield, fodder yield and their parameters in rabi sorghum hybrids (Sorghum biocolor (L.) Monech). Electron. J. Plant Breed. 7, 730–736. doi:10.5958/0975-928X.2016.00095.8
Nuruzzaman, M., Sharoni, A.M., Satoh, K., Moumeni, A., Venuprasad, R., Serraj, R., Kumar, A., Leung, H., Attia, K., Kikuchi, S., 2012a. Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from crossing Aday Selection (drought tolerant) and IR64. Mol. Genet. Genomics 287, 389–410. doi:10.1007/s00438-012-0686-8
Nuruzzaman, M., Sharoni, A.M., Satoh, K., Moumeni, A., Venuprasad, R., Serraj, R., Kumar, A., Leung, H., Attia, K., Kikuchi, S., 2012b. Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from crossing Aday Selection (drought tolerant) and IR64. Mol. Genet. Genomics 287, 389–410. doi:10.1007/s00438-012-0686-8
Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J., Rokhsar, D.S., 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551–556. doi:10.1038/nature07723
Peiffer, J.A., Flint-Garcia, S.A., Leon, N.D., McMullen, M.D., Kaeppler, S.M., Buckler, E.S., 2013. The Genetic Architecture of Maize Stalk Strength. PLOS ONE 8, e67066. doi:10.1371/journal.pone.0067066
Peng, X., Zhao, Y., Li, X., Wu, M., Chai, W., Sheng, L., Wang, Y., Dong, Q., Jiang, H., Cheng, B., 2015a. Genomewide identification, classification and analysis of NAC type gene family in maize. J. Genet. 94, 377–390. doi:10.1007/s12041-015-0526-9
Peng, X., Zhao, Y., Li, X., Wu, M., Chai, W., Sheng, L., Wang, Y., Dong, Q., Jiang, H., Cheng, B., 2015b. Genomewide identification, classification and analysis of NAC type gene family in maize. J. Genet. 94, 377–390. doi:10.1007/s12041-015-0526-9
Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., Salzberg, S.L., 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11, 1650–1667. doi:10.1038/nprot.2016.095
Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J.-L., 2012. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PLoS ONE 7, e32253. doi:10.1371/journal.pone.0032253

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. doi:10.1093/bioinformatics/btp616

Smith, C.W., Frederiksen, R.A. (Eds.), 2000. Sorghum: origin, history, technology, and production, Wiley series in crop science. Wiley, New York.

Swarts, K., Li, H., Romero Navarro, J.A., An, D., Romay, M.C., Hearne, S., Acharya, C., Glaubitz, J.C., Mitchell, S., Elshire, R.J., others, 2014. Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. Plant Genome 7.

Tran, L.-S.P., Nakashima, K., Sakuma, Y., Osakabe, Y., Qin, F., Simpson, S.D., Maruyama, K., Fujita, Y., Shinozaki, K., Yamaguchi-Shinozaki, K., 2007. Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in Arabidopsis. Plant J. 49, 46–63. doi:10.1111/j.1365-313X.2006.02932.x

Updegraff, D.M., 1969. Semimicro determination of cellulose in biological materials. Anal. Biochem. 32, 420–424.

Vargas, L., Cesarino, I., Vanholme, R., Voorend, W., Soriano Saleme, M. de L., Morreel, K., Boerjan, W., 2016. Improving total saccharification yield of Arabidopsis plants by vessel-specific complementation of caffeoyl shikimate esterase (cse) mutants. Biotechnol. Biofuels 9, 1–16. doi:10.1186/s13068-016-0551-9

Wang, D., Qin, Y., Fang, J., Yuan, S., Peng, L., Zhao, J., Li, X., 2016. A Missense Mutation in the Zinc Finger Domain of OsCESA7 Deleteriously Affects Cellulose Biosynthesis and Plant Growth in Rice. PLoS ONE 11, 1–16. doi:10.1371/journal.pone.0153993

Xiao, Y., Poovaiah, C., Coleman, H.D., 2016. Expression of Glycosyl Hydrolases in Lignocellulosic Feedstock: An Alternative for Affordable Cellulosic Ethanol Production. BioEnergy Res. 9, 1290–1304. doi:10.1007/s12155-016-9766-7

Xie, Q., Frugis, G., Colgan, D., Chua, N.-H., 2000. Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev. 14, 3024–3036.

Zhao, C., 2005. The Xylem and Phloem Transcriptomes from Secondary Tissues of the Arabidopsis Root-Hypocotyl. PLANT Physiol. 138, 803–818. doi:10.1104/pp.105.060202

Zhao, C., Avci, U., Grant, E.H., Haigler, C.H., Beers, E.P., 2007. XND1, a member of the NAC domain family in Arabidopsis thaliana, negatively regulates lignocellulose synthesis and programmed cell death in xylem: XND1 blocks tracheary element differentiation. Plant J. 53, 425–436. doi:10.1111/j.1365-313X.2007.03350.x

Zhong, R., Demura, T., Ye, Z.-H., 2006. SND1, a NAC Domain Transcription Factor, Is a Key Regulator of Secondary Wall Synthesis in Fibers of Arabidopsis. Plant Cell 18, 3158–3170. doi:10.1105/tpc.106.047399

Zhong, R., Ye, Z.-H., 2015. Secondary Cell Walls: Biosynthesis, Patterned Deposition and Transcriptional Regulation. Plant Cell Physiol. 56, 195–214.

# Appendix

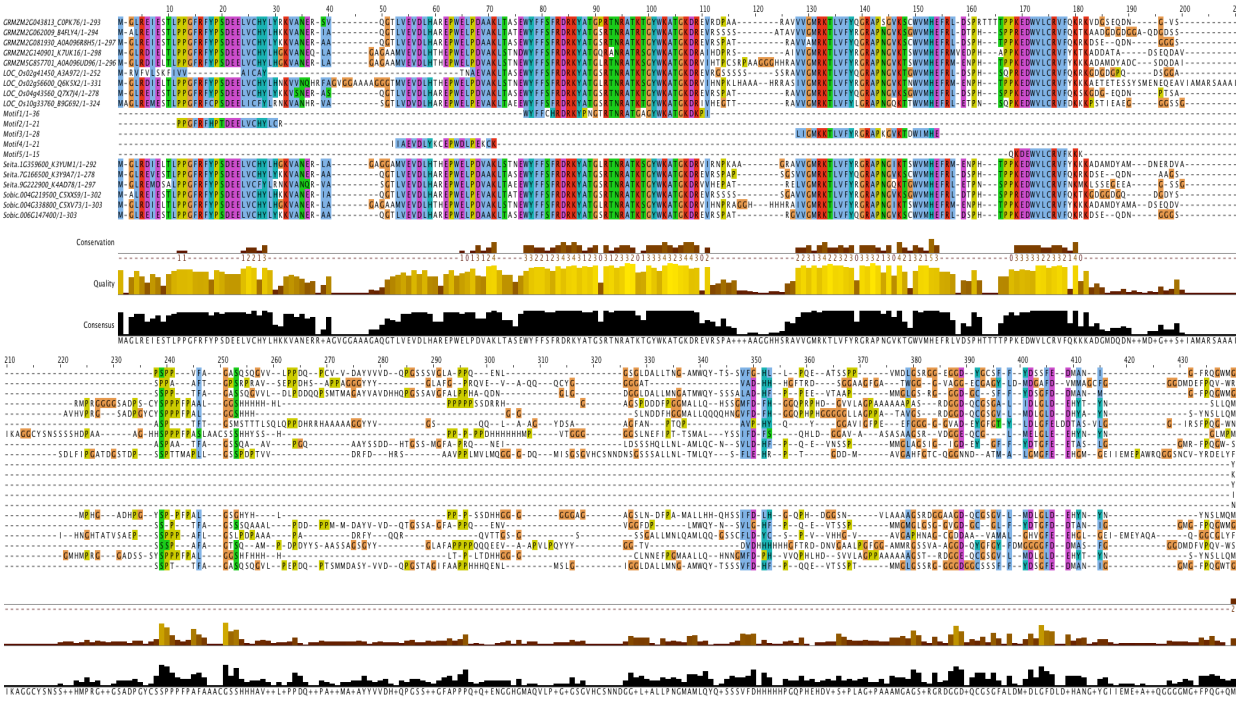

**Figure A1.** Multiple sequence alignment of monocot NAC074 peptides.

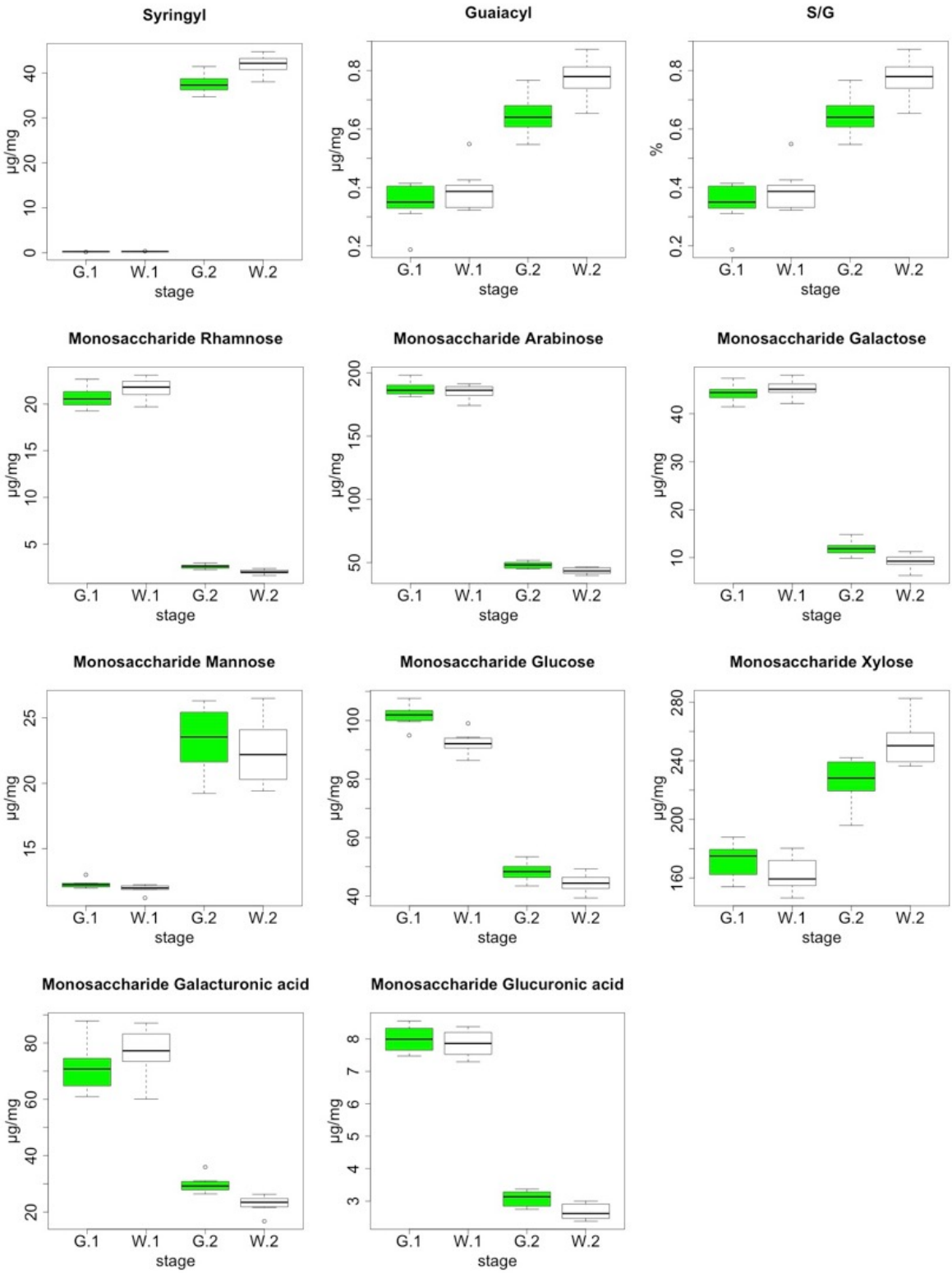
Syringyl
µg/mg
0
10
20
30
40
G.1
W.1
G.2
W.2
stage
Guaiacyl
µg/mg
0.2
0.4
0.6
0.8
G.1
W.1
G.2
W.2
stage
S/G
%
0.2
0.4
0.6
0.8
G.1
W.1
G.2
W.2
stage
Monosaccharide Rhamnose
µg/mg
5
10
15
20
G.1
W.1
G.2
W.2
stage
Monosaccharide Arabinose
µg/mg
50
100
150
200
G.1
W.1
G.2
W.2
stage
Monosaccharide Galactose
µg/mg
10
20
30
40
G.1
W.1
G.2
W.2
stage
Monosaccharide Mannose
µg/mg
15
20
25
G.1
W.1
G.2
W.2
stage
Monosaccharide Glucose
µg/mg
40
60
80
100
G.1
W.1
G.2
W.2
stage
Monosaccharide Xylose
µg/mg
160
200
240
280
G.1
W.1
G.2
W.2
stage
Monosaccharide Galacturonic acid
µg/mg
20
40
60
80
G.1
W.1
G.2
W.2
stage
Monosaccharide Glucuronic acid
µg/mg
3
4
5
6
7
8
G.1
W.1
G.2
W.2
stage

**Figure A2.** Detailed compositional analysis of *DD* and *dd* NIL internodes at the same stages shown in Figure 6. First 3 panels show lignin monomer content (syringyl, guaiacyl, and S/G ratio) and last 8 panels show monosaccharide release following TFA hydrolysis.

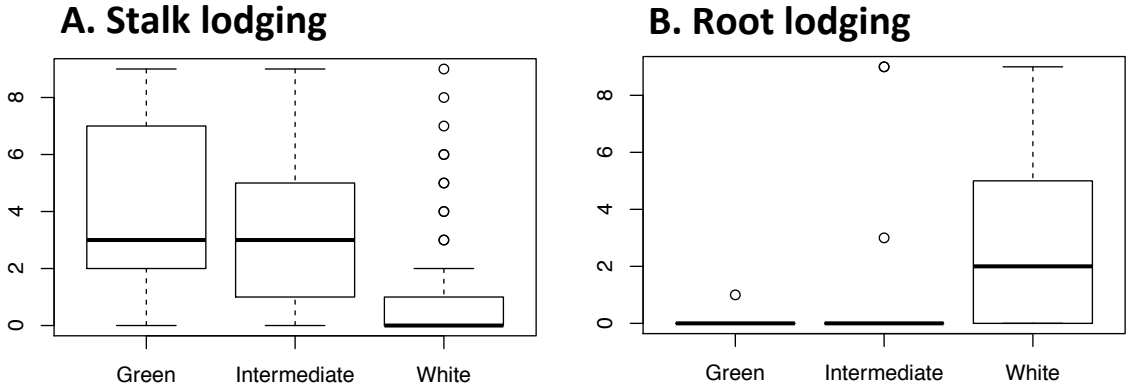


**Figure A3.** Stalk lodging (A) and root lodging (B) in diverse biomass sorghum accessions (n = 226) grouped by midrib color. Four-row, 9m$^2$ plots were scored for both types of lodging on a 0-9 scale reflecting the proportion of the plot affected. Stalk lodging predominates in green midrib (putative *dd*) accessions, whereas root lodging is more common in white midrib (putative *DD*) accessions.